A distribution-based face/nonface classification technique

Son Lam Phung
University of Wollongong, phung@uow.edu.au

Douglas Chai
Edith Cowan University

Abdesselam Bouzerdoum
Edith Cowan University, bouzer@uow.edu.au

A distribution-based face/nonface classification technique

Abstract

The core element of many existing approaches to face detection is the classification algorithm that determines if a sub-image of an input image contains a face pattern. In this paper, we present a novel and effective distribution-based face/non-face classification technique that detects frontal face patterns with possible in-plane rotation. A 15x15 input sub-image is first processed by a color filter, which verifies the presence of human skin color in the sub-image. Then, the intensity image is extracted from the identified skin color sub-image and converted into a vector in a high-dimensional space (R^{225}). Principal component analysis is employed to reduce the dimension of this space to 20. In our approach, the distributions of face and non-face patterns in the R^{20} space are modeled using mixtures of Gaussians. The parameters of the Guassian mixture models are determined through the use of the Expectation/Maximization (EM) algorithm. Finally, the classification of sub-images into face or non-face patterns is carried out through comparison of their estimated probability density functions. Experimental results have shown that the proposed technique is capable of performing highly accurate face/non-face classification.

Keywords

face, nonface, distribution, technique, classification

Disciplines

Engineering | Science and Technology Studies

Publication Details

S. Phung, D. Chai & A. Bouzerdoum, "A distribution-based face/nonface classification technique," Australian Journal of Intelligent Information Processing Systems, vol. 7, (3-4) pp. 132-138, 2002.

A Distribution-Based Face/Non-Face Classification Technique

Son Lam Phung, Douglas Chai, and Abdesselam Bouzerdoum

Edith Cowan University
School of Engineering and Mathematics
100 Joondalup Drive, Joondalup WA 6027
Perth, AUSTRALIA

Abstract

The core element of many existing approaches to face detection is the classification algorithm that determines if a sub-image of an input image contains a face pattern. In this paper, we present a novel and effective distribution-based face/non-face classification technique that detects frontal face patterns with possible in-plane rotation. A 15x15 input sub-image is first processed by a color filter, which verifies the presence of human skin color in the sub-image. Then, the intensity image is extracted from the identified skin color sub-image and converted into a vector in a high-dimensional space (R^{225}). Principal component analysis is employed to reduce the dimension of this space to 20. In our approach, the distributions of face and non-face patterns in the R^{20} space are modeled using mixtures of Gaussians. The parameters of the Gaussian mixture models are determined through the use of the Expectation/Maximization (EM) algorithm. Finally, the classification of sub-images into face or non-face patterns is carried out through comparison of their estimated probability density functions. Experimental results have shown that the proposed technique is capable of performing highly accurate face/non-face classification.

1. Introduction

The research into face detection has attracted considerable attention in recent years. It deals with the problem of detecting the presence and the location of human faces in images. Many approaches to this problem involve a scanning process in which sub-images or windows of the input image are searched exhaustively for face patterns. They use various classification algorithms to determine whether or not a sub-image contains a face pattern. For example, Rowley *et al.* [15] proposed a neural network based approach that uses multilayer perceptron to detect faces, and Feraud *et al.* [6] suggested a different neural network classifier that is based on the constrained generative model. Sung and Poggio [16], on the other hand, proposed a classifier that models the distribution of face and non-face with Gaussian clusters, found by using the elliptical *k*-means clustering technique. Osuna *et al.* [13] developed a classifier based on support vector machines, while Yang *et al.* [19] suggested a probabilistic method that uses a mixture of factor analyzers. For comprehensive reviews of these and many other face detection techniques, readers are referred to the two recent survey papers, one written by Yang *et al.* [20] and the other by Hjelmas and Low [9].

In this paper, we propose a distribution-based technique for classifying image patterns into face and non-face. The probability density functions (pdfs) of face and non-face patterns are modeled with Gaussian mixtures. The parameters of the Gaussian mixtures are determined using the EM algorithm. The paper is organized as follows. Section 2 explains the proposed face/non-face classification technique. The distribution-based approach, which uses Gaussian mixtures, to the classification task is presented in Section 3. Experimental results along with some discussion can be found in Section 4, while conclusion is given in Section 5.

2. Face/Non-Face Classification

2.1 Overview

The proposed face/non-face classification technique is to work on a digital color image. The given image is divided into non-overlapping blocks of size 15×15, and these sub-images become the input data to the classifier. The 15×15 size is chosen after considering the trade-off between classification accuracy and computation load.

This classifier is currently being developed to detect not only frontal faces but also faces with some degrees of in-plane rotation. More details can be found in Section 4. A block diagram showing the three major components, or stages, of the classification technique is depicted in Figure 1. Their functions are as follows:

- In stage 1, a color filter is applied to the input sub-image in order to detect skin color pixels. The non-existence of face pattern is flagged if no or small amount of skin-color pixels has been detected. In this case, a non-face pattern is assumed and no further processing on this sub-image is required. If, on the other hand, the sub-image contains substantial skin-color pixels then the intensity values of the sub-image is extracted and fed as input to stage 2.
- In stage 2, more pre-processing steps are performed. The 15x15 input intensity image is first histogram equalized before being converted into a column vector in $\mathbf{R}^{225}$ space. Principal component analysis (PCA) is used to reduce the dimension of this space. After applying PCA, each column vector in $\mathbf{R}^{225}$ is represented by a feature vector $\mathbf{w}$ in $\mathbf{R}^{20}$.
- In stage 3, the feature vector $\mathbf{w}$ is classified as face or non-face pattern by using a novel distribution-based algorithm. This algorithm is presented in Section 3.

As for stages 2 and 3, their respective functions are explained further in Sections 2.2 and 2.3.

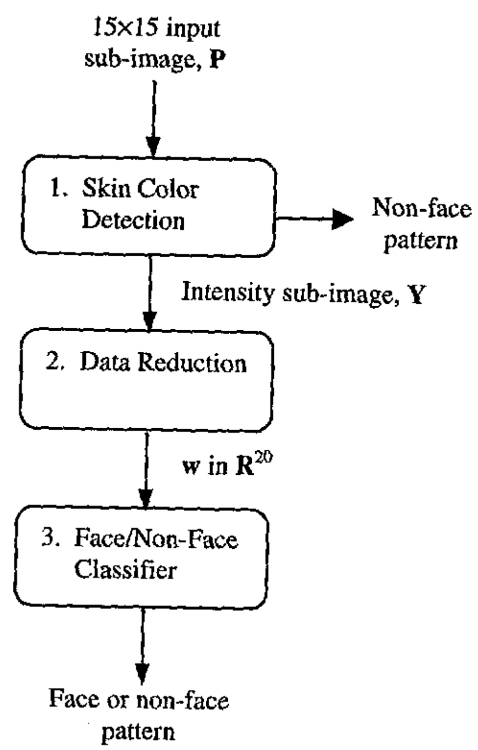

Figure 1: Block diagram of the proposed face/non-face classification technique.

2.2 Skin Color Detection

This component of the classification technique exploits the fact that human skins have distinct colors. The skin color filter used in the stage works in YCbCr color space with 8 bits per channel. The luminance value of a pixel (which corresponds to intensity) is stored in Y component, while the chrominance values are stored in Cb and Cr components. This color space is widely used in image/video coding standards such as JPEG, MPEG and H.26x. Conversions between YCbCr and RGB color spaces can be done through a linear transformation, see [4] for an example.

It has been found that skin-color region can be identified by the presence of a certain set of chrominance values that is narrowly and consistently distributed in the YCbCr color space. It has also been proven that such model is robust against different types of skin color such as white, black, yellow, brown, etc. Further details of skin color model can be found in [2,3].

Here, a simple yet effective skin color filter is used to distinguish between skin color pixel and non skin color pixel. The filter identifies a pixel i of the input sub-image as skin color if its chrominance components satisfy the following criteria:

$$Cb_i \in [75, 135] \text{ and } Cr_i \in [130, 180]. \tag{1}$$

The skin color filter partitions the input sub-image $\mathbf{P}$ into two disjoint sets: $\mathbf{P} = \mathbf{P}_{skin} \cup \mathbf{P}_{non\text{-}skin}$. The input sub-image is considered as non-face if:

$$\frac{|\mathbf{P}_{skin}|}{|\mathbf{P}|} < \tau_{skin}, \tag{2}$$

where $|\mathbf{P}_x|$ denotes the number of elements in set $\mathbf{P}_x$ and τ_{skin} is a fixed threshold.

If the sub-image $\mathbf{P}$ passes this skin color test, its corresponding intensity sub-image is extracted and sent to stage 2. Here we use the luminance values of the sub-image, which we denoted by $\mathbf{Y}$, to represent the intensity sub-image.

Note that more sophisticated filtering techniques such as those that use neural networks [14], Gaussian [12], elliptical skin model fitting [1], and mixture of Gaussians [18] can be employed instead of the model of Eq. (1).

2.3 Data Reduction

In stage 2 of our approach, the intensity sub-image $\mathbf{Y}$ is histogram-equalized so as to reduce the effects of lighting variations. From the enhanced intensity sub-image, a vector $\mathbf{x}$ of 255 elements (ie 15x15) is formed by reading $\mathbf{Y}$ column-wise.

The pattern $\mathbf{x}$ has a relatively large dimension (D = 225), and it contains some data that are not significant for classification purposes. Hence, we employ principal component analysis (PCA) to remove the irrelevant data and reduce the space dimension. In doing so, we cut down the computation load significantly. Note that to further reduce data, a mask that excludes boundary pixels from each pattern can be applied before classification – see [13, 15, 16] for further information.

PCA works as follows. Let $\{\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_N\}$ be a set of known face patterns in $\mathbf{R}^D$ space. The PCA aims to find a set of k (typically $k << D$) orthogonal vectors, $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2,\ldots, \mathbf{v}_k\}$, so that the distances (ie reconstruction errors) between $\mathbf{x}_i$ and their projections onto the subspace spanned by the k vectors are minimized. The steps for finding $\mathbf{V}$ are as follows:

- Compute the mean face vector:

$$\mathbf{x}_m = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{3}$$

- Compute the covariance matrix:

$$\mathbf{C} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \mathbf{x}_m)(\mathbf{x}_i - \mathbf{x}_m)^{\mathrm{T}} \tag{4}$$

- Compute the set of all N eigenvectors of $\mathbf{C}$. A vector $\mathbf{v}$ is an eigenvector of $\mathbf{C}$ if there exists a scalar λ such that:

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v} \tag{5}$$

  The scalar λ is called an eigenvalue of $\mathbf{C}$.

- Select k most significant eigenvectors, $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2,\ldots, \mathbf{v}_k\}$, that correspond to the largest k eigenvalues of $\mathbf{C}$.

Once the significant eigenvectors (also termed as *eigenfaces* by Turk and Pentland [17]) are computed, each new pattern $\mathbf{x}$ is represented by the projection of its deviation from the mean face vector onto the subspace spanned by the column vectors of $\mathbf{V}$:

$$\mathbf{w} = \mathbf{V}^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_m), \tag{6}$$

where $\mathbf{w}$ is a vector in k-dimensional space. In our design, we have set k to 20.

The problem of classifying the feature vector $\mathbf{w}$ into face or non-face class is dealt with in the third component of our classification technique, which is discussed in the next section.

3. Distribution-Based Classifier

The feature vector $\mathbf{w}$ for face and non-face is assumed to have arisen from two distinct, constrained but unknown distributions in $\mathbf{R}^k$ space. Let $p(\mathbf{w}|face)$ and $p(\mathbf{w}|nonface)$ be the pdfs of face and non-face patterns, respectively. By applying the Bayes decision rule for minimum cost [8,4] to this two-class (ie face and non-face) classification problem, we classify the feature vector $\mathbf{w}$ as face if the following equation is satisfied:

$$\frac{p(\mathbf{w}\mid face)}{p(\mathbf{w}\mid nonface)} \geq \tau_p, \tag{7}$$

where τ_p is a fixed threshold, which is dependent on various classification costs and *a priori* probabilities of face and non-face. However, the value of τ_p is often determined experimentally. The derivation of (7) is given in the Appendix.

A difficult problem remains, and that is to find the probability density functions $p(\mathbf{w}|face)$ and $p(\mathbf{w}|nonface)$. The common non-parametric approaches to pdf estimation such as histograms and kernel-based methods are not feasible in this situation due to the high dimension of the feature vector. We, therefore, propose a parametric approach for estimating $p(\mathbf{w}|face)$ and $p(\mathbf{w}|nonface)$ using mixtures of Gaussians.

3.1 Gaussian Mixture Modeling of Face PDF

The Gaussian mixture technique models the pdf of face pattern $\mathbf{w}$ as a linear combination of a set of G Gaussian components:

$$p(\mathbf{w}\mid face) = p(\mathbf{w}) = \sum_{i=1}^{G}\pi_i g(\mathbf{w};\theta_i), \tag{8}$$

where π_i and θ_i are the mixing factor and the parameter of the i-th component, respectively. The mixing factor must satisfy the following conditions:

$$\sum_{i=1}^{G}\pi_i = 1 \quad \text{and} \quad \pi_i \geq 0, i = \overline{1,G}. \tag{9}$$

In our case, each component $g(\mathbf{w};\theta_i)$ is a Gaussian, which is characterized by a mean vector μ_i and a covariance matrix Σ_i. Hence,

$$\theta_i = \{\mu_i, \Sigma_i\},$$

$$g(\mathbf{w};\mu_i,\Sigma_i) = (2\pi)^{-\frac{k}{2}}|\Sigma_i|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{w}-\mu_i)^T\Sigma_i^{-1}(\mathbf{w}-\mu_i)\}. \tag{10}$$

Suppose we have a training set of face vectors $\{\mathbf{w}_1, \mathbf{w}_2,\cdots, \mathbf{w}_N\}$. The parameter set of the model $\phi = \{\pi_i, \theta_i\}$ for $i = \overline{1,G}$ is estimated from the training vectors, and it must satisfy a condition known as *maximum-likelihood,* which requires the following joint probability of occurrence of the training vectors to be maximized:

$$\mathrm{L} = \prod_{i=1}^{N}p(\mathbf{w}_i). \tag{11}$$

A common method for parameter estimation using maximum-likelihood is the Expectation/Maximization (EM) algorithm.

3.1.1 EM Algorithm

The EM algorithm [5,7] starts with an initial estimate of the parameter set ϕ. This estimate is used to compute the pdf of $\mathbf{x}$ as in (8). The pdf of $\mathbf{x}$ is then used to compute a revised estimate of ϕ (ie mixing factors π_i, means μ_i, and covariance matrices Σ_i). This process continues for a fixed number of iterations until there is little change in ϕ, or until the following log-likelihood function exceeds a certain threshold:

$$\ln \mathrm{L} = \sum_{i=1}^{N}\ln p(\mathbf{w}_i). \tag{12}$$

Given that the probability of sample $\mathbf{w}_j$ belonging to the i-th component is defined as

$$p^t(i\mid\mathbf{w}_j) = \frac{\pi_i^t g(\mathbf{w}_j;\mu_i^t,\Sigma_i^t)}{p^t(\mathbf{w}_j)}, \tag{13}$$

where superscript t denotes the iteration number, the revised estimate of ϕ at each stage is provided as follows:

- The revised estimate of the mixing factor for the i-th component is computed by

$$\pi_i^{t+1} = \frac{1}{N}\sum_{j=1}^{N}p^t(i\mid\mathbf{w}_j) = \frac{P}{N}. \tag{14}$$

- The revised estimate of the mean for the i-th component is calculated by

$$\mu_i^{t+1} = \frac{1}{P}\sum_{j=1}^{N}p^t(i\mid\mathbf{w}_j)\mathbf{w}_j. \tag{15}$$

- The revised estimate of the covariance matrix for the i-th component is given by

$$\Sigma_i^{t+1} = \frac{1}{P}\sum_{j=1}^{N}p^t(i\mid\mathbf{w}_j)(\mathbf{w}_j-\mu_i^t)(\mathbf{w}_j-\mu_i^t)^{\mathrm{T}}. \tag{16}$$

3.1.2 EM Initialization

The initial estimate for the EM algorithm can influence its convergence speed and final result. In our approach, Kohonen's Self-Organizing Map (SOM) algorithm [10] is applied to divide the training vector set $\{\mathbf{w}_j\}$ for $j = \overline{1,N}$ into G clusters, C_i, for $i = \overline{1,G}$. The initial estimate of the parameter set ϕ is then determined as follows:

$$\pi_i^0 = \frac{|C_i|}{N}, \tag{17}$$

$$\mu_i^0 = \mathop{\mathrm{E}}_{\mathbf{w}_j\in C_i}\{\mathbf{w}_j\}, \tag{18}$$

$$\Sigma_i^0 = \mathop{\mathrm{E}}_{\mathbf{w}_j\in C_i}\{(\mathbf{w}_j-\mu_i^0)(\mathbf{w}_j-\mu_i^0)^{\mathrm{T}}\}, \tag{19}$$

where $|C_i|$ is the size of cluster C_i and $\mathrm{E}\{\}$ is the expectation operator.

3.2 Gaussian Mixture Modeling of Non-Face PDF

The pdf for non-face $p(\mathbf{w}|nonface)$ is also modeled as a mixture of Gaussians using essentially the same steps as described in Section 3.2. From a large number of non-face images, a set of non-face feature vectors $\mathbf{w}$ is obtained. This set is clustered using Kohonen's SOM algorithm to obtain an initial estimate of the parameters of the Gaussian mixture model. The EM algorithm is then performed to obtain the final estimate of the parameters. Note that the number of Gaussian components for non-face does not need to be the same as in $p(\mathbf{w}|face)$. Compared to the Gaussian mixture model for face, the model for non-face needs to be updated more frequently as new non-face patterns are presented to the classifier. This is due to the difficulty in finding a representative non-face pattern.

3.3 Face/Non-Face Classification

Once the parameters of the Gaussian mixture models are determined, face/non-face classification of the feature vector is carried out as follows. A feature vector is classified as face if it satisfies two conditions:

$$\text{Condition 1: } \frac{p(\mathbf{w}\mid face)}{p(\mathbf{w}\mid nonface)} > \tau_p \tag{20}$$

and

$$\text{Condition 2: } p(\mathbf{w}\mid face) > \tau_{face} \tag{21}$$

The first condition is the Bayes decision rule for minimum cost as mentioned earlier. The second condition is to help remove non-face patterns that pass the first condition.

4. Experiment Results

4.1 Training Procedure

The training set of face vectors was generated from the AR face database [11]. This database consists of more than 4,000 color frontal-face images of 126 people (70 men, 56 women) with various facial expressions that range from neutral, smile to anger. The images are of size 768×576. From this database, we extracted 1,000 face images, each of which has a size of 15×15. To generate sufficient training data as well as to detect face patterns that are influenced by in-plane rotation, each original face image was rotated by angles of $\pm i \times 5°$ for $i = \overline{1,10}$ to give extra 20 face images. The training set consisted of 21,000 face vectors.

PCA was performed on 1,000 face vectors and a set of $k = 20$ most significant eigenvectors was found. Each input pattern $\mathbf{x}$ was then represented by a feature vector $\mathbf{w}$ as computed in (6).

The Gaussian mixture model for face in R^{20} space consists of $G = 5$ components. The entire training set of 21,000 feature vectors $\mathbf{w}$ was used to develop the model. The Gaussian mixture model for non-face also consists of $G = 5$ components. A set of 21,000 non-face images randomly chosen from 100 training images was used to construct the non-face pdf. The various parameters of the face/non-face classifier are summarized in Table I.

Table I: Parameters of the face/non-face classifier.

Description	*Value*
Input sub-image with frontal in-plane rotation	Size: 15×15 Format: YCbCr
Skin color range	$Cb_i \in [75, 135]$ $Cr_i \in [130, 180]$
Skin color test threshold	$\tau_{skin} = 0.5$
No. of eigenfaces	$k = 20$
No. of Gaussian components	$G = 5$ each
No. of face vectors for training	$N = 21{,}000$
No. of non-face vectors for training	$N = 21{,}000$
Max. in-plane rotation allowed	$\theta = \pm 50°$

4.2 Testing Procedure

To overcome the difficulty of finding representative non-face images and also a test set that consists of a representative mixture of face and non-face images, the following testing strategy was adopted. The face/non-face classifier was evaluated on two different test sets. Test set 1 consisted of only face patterns and was used to estimate the correct detection rate. Test set 2 consisted of only non-face patterns and was used to estimate the correct rejection rate.

Test set 1 consisted of 650 color images. These images of size 15×15 were obtained from various sources including WWW and TV programs. The face images were mostly frontal with possible in-plane rotation and of people from diverse ethnicity ranging from Asian, African to European. Test set 2 consisted of 184,000 non-faces images. These non-face patterns were obtained by using a program that randomly selected images of size 15×15 from 200 color images of various sizes. These color images, sized between 10^5 to 10^6 pixels, were obtained from both natural and man-made landscape photographs that contained no face patterns.

4.3 Results and Discussion

The classification results on the two test sets are provided in Table II.

They show that the proposed face/non-face classifier is capable of performing highly accurate classification. Most false rejections occur when the face is taken under views that are markedly different from the frontal views or under extreme lighting condition. A set of representative results of correct classification (ie correct detection and rejection) are illustrated in Figures 2 and 3. Note that all samples of correct rejection as shown in Figure 3 have skin-color appearance and therefore they were not picked up as non-face pattern in stage 1 of the classification; however they were detected as non-face in stage 3 of the classification process.

Table II: Face/non-face classifier testing results.

Test set 1	650 faces
Test set 2	184,000 non-faces
Correct detection rate	89.2%
Correct rejection rate	98.9%

Figure 1: Samples of correct detection.

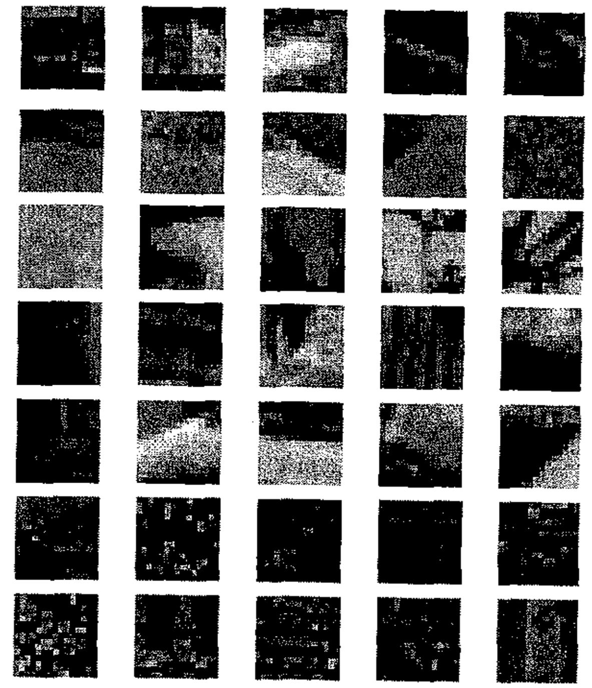

Figure 2: Samples of correct rejection.

Our proposed face/non-face classification technique can accept gray-scale input images as the skin color detection stage can be by-passed. Therefore, the proposed classification scheme is not restricted to only color input images. The main reason for inclusion of skin color detection at the beginning of the classification scheme is to speed up the whole process, as sub-images with non skin color characteristics can be reliably and quickly identified as non-face patterns without the need of further testing.

5. Conclusion

This paper presents a novel and accurate distribution-based classification technique for face and non-face patterns. The correct detection rate of 89.2% and the correct rejection rate of 98.9% were achieved. The classifier involves a skin color filter and PCA for dimension reduction of the input space. The key element of the classifier is the parametric estimation of probability density functions for face and non-face patterns using mixtures of Gaussians. The parameters of the estimation are found by using the EM algorithm. Face and non-face patterns are then classified by comparing the probability density functions.

The current classifier caters for the detection of frontal faces with some degree of in-plane rotation. We believe that the proposed approach is readily extendable to faces under other different viewing angles. A possible method is to obtain a more comprehensive training set that covers greater viewing angles, and reconstruct the Gaussian models. This will be carried out in our future research work.

6. Acknowledgment

The authors wish to express sincere thanks to Dr Martinez at Purdue University for providing the AR face image database. The authors are also grateful to Mr Tivive at Edith Cowan University for taking part in preparing the images used in this work.

7. Appendix

For clarity, let C_1 and C_2 denote face class and non-face class, respectively. Let $P(C_1|\mathbf{w})$ and $P(C_2|\mathbf{w})$ be the *a posteriori* probabilities of C_1 and C_2, respectively, (*i.e.*, probability of an observed pattern $\mathbf{w}$ belonging to a class). Let c_{ij} for i,j = {1,2} denotes the cost of classifying a pattern $\mathbf{w}$ into class C_i while it actually belongs to class C_j.

The cost of classifying a pattern $\mathbf{w}$ into class C_1 is given by

$$R_1(\mathbf{w}) = c_{11}.P(C_1|\mathbf{w}) + c_{12}.P(C_2|\mathbf{w}). \tag{22}$$

Similarly, the cost of classifying a pattern $\mathbf{w}$ into class C_2 is given by

$$R_2(\mathbf{w}) = c_{21}.P(C_1|\mathbf{w}) + c_{22}.P(C_2|\mathbf{w}). \tag{23}$$

The pattern $\mathbf{w}$ will be classified into the class that gives the minimum classification cost:

$$\begin{cases} R_1(\mathbf{w}) \geq R_2(\mathbf{w}) \Rightarrow \mathbf{w} \in C_1 \\ R_1(\mathbf{w}) < R_2(\mathbf{w}) \Rightarrow \mathbf{w} \in C_2 \end{cases}. \tag{24}$$

By combining with (22) and (23), this decision rule can be rewritten as follows:

$$\begin{cases} \dfrac{P(C_1 \mid \mathbf{w})}{P(C_2 \mid \mathbf{w})} \geq \dfrac{c_{22} - c_{12}}{c_{11} - c_{21}} \Rightarrow \mathbf{w} \in C_1 \\ \dfrac{P(C_1 \mid \mathbf{w})}{P(C_2 \mid \mathbf{w})} < \dfrac{c_{22} - c_{12}}{c_{11} - c_{21}} \Rightarrow \mathbf{w} \in C_2 \end{cases} \tag{25}$$

According to Bayes theorem [8], the *a posteriori* probabilities $P(C_i|\mathbf{w})$ can be expressed in terms of *a priori* probabilities $P(C_i)$, class-conditional densities $p(\mathbf{w}|C_i)$ and unconditional probability $p(\mathbf{w})$:

$$P(C_1 \mid \mathbf{w}) = \frac{p(\mathbf{w} \mid C_1)P(C_1)}{p(\mathbf{w})}, \tag{26}$$

$$P(C_2 \mid \mathbf{w}) = \frac{p(\mathbf{w} \mid C_2)P(C_2)}{p(\mathbf{w})}. \tag{27}$$

The decision rule in (25) becomes

$$\begin{cases} \dfrac{p(\mathbf{w} \mid C_1)}{p(\mathbf{w} \mid C_2)} \geq \tau_p & \Rightarrow \mathbf{w} \in C_1 \\ \dfrac{p(\mathbf{w} \mid C_1)}{p(\mathbf{w} \mid C_2)} < \tau_p & \Rightarrow \mathbf{w} \in C_2 \end{cases}, \quad (28)$$

where τ_p is a threshold that depends on the classification costs and *a priori* probabilities of face and non-face patterns, and it is defined as

$$\tau_p = \frac{c_{22} - c_{12}}{c_{11} - c_{21}} \cdot \frac{P(C_2)}{P(C_1)}. \quad (29)$$

8. References

[1] Bojic, N. and Pang, K. K. "Adaptive skin segmentation for head and shoulder video sequence," *SPIE Conference on Visual Communication and Image Processing (VCIP'2000)*, Perth, Australia, vol. 4067, part 2, pp.704-711, Jun. 2000.

[2] Chai, D. and Ngan, K. N., "Locating facial region of a head-and-shoulders color image," *IEEE International Conference on Automatic Face and Gesture Recognition (FG'98)*, Nara, Japan, pp. 124-129, Apr. 1998.

[3] Chai, D. and Ngan, K. N., "Face segmentation using skin color map in videophone applications," *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, vol. 9, pp. 551-564, Jun. 1999.

[4] Chai, D. and Bouzerdoum, A., "A Bayesian approach to skin color classification in YCbCr color space," *IEEE Region Ten Conference* (TENCON'2000), Kuala Lumpur, Malaysia, vol. II, pp. 421-424, Sep. 2000.

[5] Dempster, A. P., Laird, N. M. and Rubin, D. B., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, pp. 1-38, 1977.

[6] Feraud, R., Bernier, O. J., Viallet, J.-E. and Collobert, M., "A fast and accurate face detector based on neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 42-53, Jan. 2001.

[7] Figueiredo, M. A. T. and Jain, A. K., "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381-396, Mar. 2002.

[8] Fukunaga, K., *Introduction to statistical pattern recognition*, Boston: Academic Press, 2nd edition 1990.

[9] Hjelmas, E. and Low, B. K., "Face detection: a survey," *Computer Vision and Image Understanding*, pp. 236-274, 2001.

[10] Kohonen, T., *Self-Organizing Maps*, Berlin: Springer, 2nd edition, 1997.

[11] Martinez, A. M. and Benavente, R., "The AR Face Database," *CVC Technical Report No. 24*, School of Electrical & Computer Engineering, Purdue University, Jun. 1998.

[12] Menser, B. and Wien, M., "Segmentation and tracking of facial regions in color image sequences," *SPIE Conference on Visual Communication and Image Processing (VCIP'2000)*, Perth, Australia, vol. 4067, part 2, pp. 731- 740, Jun. 2000.

[13] Osuna, E., Freund, R. and Girosi, F., "Training support vector machines: an application to face detection," *Computer Vision and Pattern Recognition*, Puerto Rico, June 1997.

[14] Phung, S. L., Chai, D. and Bouzerdoum, A., "A universal and robust human skin color model using neural networks," *INNS-IEEE International Joint Conference on Neural Networks*, Washington DC, USA, Jul. 2001.

[15] Rowley, H. A., Baluja, S. and Kanade T., "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 23-38, Jan. 1998.

[16] Sung, K. K. and Poggio, T., "Example-based learning for view-based human face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 39-51, Jan. 1998.

[17] Turk, M. A. and Pentland, A. P., "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-96, 1991.

[18] Yang, M.-H. and Ahuja, N., "Gaussian mixture model for human skin color and its applications in image and video databases," *SPIE Conference on Storage and Retrieval for Image and Video Databases*, San Jose, USA, Jan. 1999.

[19] Yang, M.-H., Ahuja, N. and Kriegman, D., "Face detection using a mixture of factor analyzers," *IEEE International Conference on Image Processing (ICIP'99)*, Kobe, Japan, 1999.

[20] Yang, M.-H., Kriegman, D. and Ahuja, N., "Detecting faces in images: a survey," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34-58, Jan. 2002.